

Parte 2 – Reconhecimento de Fala

1. Breve Introdução

Esta parte do trabalho laboratorial tem como objetivo, simular um sistema automático de reconhecimento de palavras isoladas com adaptação ao orador através da ferramenta **HTK**. Com esta ferramenta é possível construir modelos de *Markov* não observáveis, permitindo gerar reconhecedores no âmbito de processamento da fala. Ao longo desta pequena demonstração de resultados, as nossas tabelas devem ser suportadas pelas imagens na directoria “**Anexos**”, que comprovam a veracidade dos valores obtidos.

2. Taxa de Sucesso para os modelos com orador independente e speaker adaptado, considerando apenas letras e números

Na primeira fase deste trabalho foram executados os scripts da baseline sem qualquer modificação. A sequência de instruções para executar a baseline encontra-se no script **do_baseline.sh**. Desta forma podemos correr a baseline com a instrução `./scripts/do_baseline.sh` na raiz do tool_grid. Os resultados do reconhecimento encontram-se em `results/testSA_train3mix/testSA_train3mix.txt`. As instruções seguintes (1-6) ilustram o conteúdo da nossa baseline.

- 1) `./scripts/build_flists_train.sh`
- 2) `./scripts/build_flists_test.sh`
- 3) `./scripts/do_mfcc_train.sh`
- 4) `./scripts/do_mfcc_test.sh`
- 5) `./scripts/do_train.sh train3mix mfcc features/mfcc/train`
- 6) `./scripts/do_recog.sh mfcc train3mix testSA features/mfcc/test/`

Alterando o script **do_recog.sh** nos locais apropriados (estão devidamente comentados no próprio script), passámos para um modelo de orador independente, passando a utilizar esse mesmo modelo para todos os **34 speakers**. Modelo esse que se encontra em `models/train/SI/`.

- 7) `./scripts/do_recog.sh mfcc train3mix testSI features/mfcc/test/`

Através da execução da **instrução 7** obtivemos as seguintes taxas de sucesso para **MIX de 3**:

Modelo de Teste	Percentagem [%]
Adaptado	93,2758
Independente	83,7875

Tabela 1 – Taxas de sucesso para os modelos de “test”

3. Melhorar a baseline e repetir o processo da alínea 2

Nesta fase do trabalho, para melhorar a taxa de sucesso do reconhecedor, programou-se o script **do_allMixturesAndSpeakerType.sh** para se poder analisar as resultantes taxas de sucesso que provêm dessas alterações. Os parâmetros testados pelo script incluem 5 e 7 misturas de Hidden Markov Model, e variar o orador de adaptativo para independente. As taxas de sucesso para todos os casos testados podem ser observadas na tabela abaixo:

Modelo de Teste	Percentagem [%]	
	5 Mix	7 Mix
Adaptado	94,2364	94,6971%
Independente	87,5711	90,1000%

Tabela 2 – Taxas de sucesso para os modelos de “test” com alterações na Baseline

Os ficheiros dos resultados para cada tipo de reconhecimento encontram-se em **"results/testXX_trainYmix/testXX_trainYmix.txt"**, onde XX toma valores SA ou SI para oradores (speakers) adaptados ou independentes. O valor Y pode ser 3, 5 ou 7 para o número de misturas.

4. Fazendo agora a adaptação para os novos domínio e vocabulário

Nesta fase do trabalho, foi alterada a gramática e o dicionário para reconhecer matrículas no formato **"número número letra letra número número"**. Estas alterações foram feitas nos ficheiros grammar2, dict2 e wlist2 na directoria **"etc/"**. Por cada elemento do grupo foram gravadas 10 matrículas diferentes que se encontram em **"data/test2/idZ/"**. Tendo Z o valor 1 ou 2 consoante o elemento do grupo que fez as gravações. Os ficheiros que resultaram do reconhecimento para estas gravações encontram-se em **"results/plateSI_train3mix/plateSI_train3mix.txt"**. Neste caso, os resultados são apenas considerados no contexto de orador independente uma vez que nenhum dos alunos faz parte dos oradores do treino.

Modelo de Teste	Percentagem [%]		
	3 Mix	5 Mix	7 Mix
Independente	14,2857	11,9048	21,4286

Tabela 3 - Taxas de sucesso para os ficheiros de “test” gravados pelo grupo

5. Anotação automática de palavras para os novos ficheiros de teste

Para concluir esta segunda parte do trabalho laboratorial, foi executado o reconhecedor em alinhamento forçado. Para a resolução desta alínea foi programado o script **do_simpleForceAlign.sh** para orador independente e foi escolhido um **MIX de 7**, pois foi o que apresentou melhor taxa de sucesso na **alínea 4**. Tal como foi referido na alínea anterior, cada elemento do grupo gravou 10 matrículas diferentes. Os ficheiros das labels estão em **"labels/plates/"**. Os ficheiros que resultaram podem ser encontrados na directoria **"forced_alignments/"**.

Para podermos analisar os resultados destes alinhamentos, foi criado o script **mlfToLab.sh** que remove o conteúdo que o wavesurfer não reconhece e retorna todos os ficheiros presentes no *"forced_alignments/"* em *"forced_alignmentsLAB/"*.

Através do wavesurfer é possível analisar os ficheiros com as suas transcrições, podemos ainda inferir que o **HVite** desempenha um alinhamento bastante correto das labels.

6. Decisões e Comentários Finais